PENNSYLVANIA STATE LIBRARY
DOCUMENTS SECTION

STATEWIDE EPIDEMIOLOGIC SURVEILLANCE: CONCEPT AND METHODOLOGY

George K. Tokuhata, Dr.P.H., Ph.D.

Director

Division of Epidemiology Research

Pennsylvania Department of Health

July 1990

PHE 68.2 S797e c.1
Tokuhata, George Kazunari,
1925-
Statewide epidemiologic
surveillance

STATEWIDE EPIDEMIOLOGIC SURVEILLANCE: CONCEPT AND METHODOLOGY

Table of Contents

		Pages
I.	Introduction	1
II.	Epidemiologic Surveillance	2
III.	Objectives:	
	A. General Objectives	6
	B. Specific Objectives	7
IV.	Methods:	
	A. Geographic Areas and Populations	8
	B. Data Sources and Health Indices:	
	(1) Population Data	11
	(2) Mortality-Natality Data	12
	(3) Cancer Registry Data	13
	(4) Minimum Data Set	14
	(5) Health Indices	15
	C. Data Analysis:	
	(1) Adjustment (Standardization) of Health Indices	16
	(2) Modes of Comparison:	
	(a) Cross-Sectional Comparison Between Study and Control Areas	20
	(b) Temporal Comparison Within Study Population	21
	(c) Temporal-Spatial Comparison	22
	D. Surveillance Sequence	22
	E. Limitations and Qualifications	24

V. Operational Procedures 26

VI. Significance and Public Health Implications 28

References 31

STATEWIDE EPIDEMIOLOGIC SURVEILLANCE: CONCEPT AND METHODOLOGY

I. Introduction:

The need for epidemiologic surveillance in populations where potential environmental health risks exist has long been recognized (1-7). In 1974, an anti-nuclear activist alleged that releases of radioactive materials from the Shippingport Atomic Power Station (Beaver County, Pennsylvania) caused increases in certain diseases and mortality in surrounding populations. As a result, a Governor's Fact-Finding Committee was established to investigate this allegation. When the results of its investigation were reported, the Committee recommended that a systematic and routine epidemiologic surveillance program around nuclear power plants be developed using the readily available vital statistics data. The Committee also recommended that the Pennsylvania Department of Health develop and implement a Statewide Cancer Registry so that the incidence of cancer could be incorporated into the already existing vital statistics data set for surveillance purpose.

More recently, formal recommendations were made in a Pennsylvania State Health Plan to design and implement an improved epidemiologic surveillance system for diseases and deaths of occupational, industrial and environmental etiology. The nuclear accident at Three Mile Island (TMI) on March 28, 1979 further provided an immediate impetus for the development of a systematic method of monitoring or evaluating the health status of a local population in the vicinity of a nuclear power plant.

Digitized by the Internet Archive
in 2017 with funding from
This project is made possible by a grant from the Institute of Museum and Library Services as administered by the Pennsylvania Department of Education through the Office of Commonwealth Libraries

https://archive.org/details/statewideepidemi00toku

"Epidemiologic surveillance" now represents one of the important priorities for the State Health Department.

Subsequent to the TMI nuclear accident, the Pennsylvania Department of Health initiated a major effort in the planning and development of a baseline data system, which would be necessary for such epidemiologic surveillance in the Commonwealth. This initiative included the early development of the mortality and natality surveillance system by minor civil divisions (cities, boroughs, and townships). The Statewide MCD-based mortality and natality surveillance system became the conceptual and methodological foundation upon which the epidemiologic surveillance around nuclear power plants or any other potentially high-risk areas could be built, where certain selected MCDs are grouped together for systematic and continuous observation.

II. Epidemiologic Surveillance:

The purpose of this report is to re-define the concept and methodology of "epidemiologic surveillance" in a broader and more generalized context so that the system, when developed and implemented, can serve similar, but mutiple objectives within the Commonwealth. Specifically, such a system should be capable of monitoring potential health problems among local residents in a delineated geographic area or community that may be associated with the existing nuclear power plant, toxic waste dump site, industrial plant polluting the environment (air, water, soil), or in a community where no apparent environmental hazards are

recognized. (Diseases for which causes are known, such as infectious/communicable diseases are not considered in this context.)

Before describing the Statewide Epidemiologic Surveillance System in technical detail, several important conceptual characteristics of the System should be addressed.

First, the major function of "surveillance" is to monitor or detect any potentially "abnormal" (significantly different from expectation or norm) trend or pattern of diseases or deaths in a population.

Second, the surveillance system is designed as a "screening," rather than "diagnostic" device; that is, significant changes and/or differences in health indices (diseases or deaths), if found, based on the surveillance data will be regarded only as "symptomatic signs" or "potential health problems." There are a myriad of environmental, occupational, social, demographic and lifestyle factors that can affect the health status of a population.

If and when potentially important changes and/or differences were noted in a population, the need for a further "diagnostic" study would be determined, and if warranted, a detailed epidemiologic study would be developed in which new epidemiologic data are collected to identify possible cause(s). In other words, this surveillance system, by itself, is not intended to test any specific hypotheses or to determine health effects of any given causative agents, but rather is designed to

determine whether or not potential health problems (important changes and/or differences in any specified health indicators) exist within a target population, regardless of causes, that can be assessed with the already available epidemiologic data.

Third, the basic intent of the surveillance system of this nature is to establish an on-going and automated mechanism of compiling, analyzing, and updating relevant and readily available health data according to a standardized geographic unit throughout the Commonwealth. Such a system can serve multiple surveillance purposes described earlier.

Fourth, the surveillance system being described in this report would serve as a vehicle (means) for public education as well as political action regarding issues related to perceived versus real health problems for any given potentially "high-risk" areas within the State. While the general public can be informed of a perceived health problem without supporting epidemiologic evidence, the identification of a real health problem, if and when documented, can be used for a further epidemiologic study and/or for an appropriate political action, including remedial recommendations to concerned Federal and/or State agencies. The skeptical general public, who often are the recipients of biased or distorted information from specific interest groups, need to be properly informed.

Fifth, the surveillance system being described here should also be conceived and structured in a larger, interagency functional context. For example, the Statewide Epidemiologic

Surveillance Around Nuclear Power Plants conducted by the State Health Department complements the existing program efforts of other State agencies, such as DER's radiation monitoring and the Department of Agriculture's food, milk, and animal monitoring. These efforts are consistent with the Commonwealth's Radiological Emergency Plans as well as U.S. Nuclear Regulatory Commission's mandate and the objectives of the U.S. Environmental Protection Agency and the Federal Emergency Management Agency. Likewise, the Statewide surveillance system oriented toward areas where potential health risks exist from toxic chemical substances needs to be coordinated with the responsibilities of such agencies as the Pennsylvania Department of Environmental Resources, the U.S. Environmental Protection Agency, and the Agency for Toxic Substance and Disease Registry of the U.S. Centers for Disease Control.

Sixth, the surveillance system being described would also be able to provide baseline health data in a timely manner, which would be invaluable in the event of a nuclear or other industrial accident when unscheduled releases of a variety of toxic substances can occur in a selected community.

III. Objectives:

A. General Objectives:

One of the basic duties of the State Health Department is to protect the health of the people of the Commonwealth and to determine and employ the most efficient and practical means for the prevention and suppression of disease. Inherent in this broad duty is a knowledge of the health status of the people and of the determinants of the diseases which afflict them.

The ultimate objective of epidemiologic surveillance then, is the timely provision, analysis and interpretation of morbidity, mortality, natality and demographic data needed to evaluate the health status of designated populations in Pennsylvania. Baseline health data, together with relevant environmental and other mitigating factors will facilitate the development of the means of preventing or suppressing diseases and health problems in the designated populations.

However, there is no single comprehensive surveillance system that can meet all such needs. Certain differences do exist in the methodology of disease surveillance between those for which the etiology is known (e.g., infectious/communicable diseases) and those for which the etiology is unknown (e.g., chronic diseases).

B. Specific Ojectives:

In view of practical considerations and the availability of needed data and other resources, the specific objectives of the Statewide Epidemiologic Surveillance System are:

(1) To develop and implement a health data system which will insure the continuous and systematic compilation, analysis and interpretation of the available relevant data according to grouped minor civil divisions (MCDs) throughout the Commonwealth.

(2) To devise an appropriate method of grouping a certain number of adjacent MCDs considered to be sharing some common, potential health risks associated with the existing environmental hazards.

(3) To devise a method of identifying/selecting an appropriate number of MCDs (control area) in the Commonwealth that can be used for comparison with a group of MCDs already delineated for a particular study/investigation.

(4) To establish baseline health indices (minimum required) based on MCDs and update such indices for subsequent years for review and analysis.

(5) To monitor and detect any apparently significant changes over time (within the same selected area) and/or significant differences from norms (compared with the control area) that may have occurred or been observed in the study area.

(6) To determine if further studies should be initiated to confirm such changes or differences in health indicators, if and when observed. If significant changes or differences were

confirmed, an in-depth epidemiologic study would be conducted to determine whether such changes or differences are related to living in the area under study or to other socio-demographic factors. At that stage of investigation, possible linkage would be examined between the observed, elevated health indices and the documented, unusual environmental exposures/releases, if any.

IV. Methods:

A. Geographic Areas and Populations:

One of the basic considerations in the development of an epidemiolgic surveillance system was the delineation of geographic units to be included in the system. These areas should not be so large that the health influences of any of the identified, potential risk sources would be too diluted to be meaningful. On the other hand, the areas should not be too small to yield an epidemiologically useful result, i.e., the population size should not be too small to yield reliable statistics.

Because of these considerations, caution must be exercised in grouping any number of adjacent MCDs to be designated as a study or control (comparison) area. Grouping of MCDs as a study area may also be influenced by other conditions, such as the availability of monitored environmental data that would eventually be useful if and when the investigation reaches this stage. Also important to consider is the variability in the geographic spread and population density of each MCD in the Commonwealth.

Regardless of "study" or "control" area delineation, the most important consideration for the purpose of surveillance and for an in-depth epidemiologic study, when conducted, is the existing, and often serious residence reporting bias according to the selected MCDs under study. Specifically, births and deaths as well as other health-related events are over-represented for those who live in cities and boroughs where U.S. post offices (thus, mailing addresses) are located. Contrarily, these events are under-represented for those who live in townships where no U.S. post offices are located. This inherent problem of mailing versus residential addresses often results in an unreasonably inflated disease/death rate for selected boroughs and cities, while an unreasonably deflated disease/death rate is attributed to selected townships.

Because of the multi-purpose surveillance system under consideration, no single geographic area consisting of a given number of MCDs can be used as a common base or "control" area for all comparisons. Therefore, some systematic method must be devised to group an appropriate number of suitable MCDs for each "study" area. Any combination of available MCDs in the Commonwealth which meet certain specific criteria can be designated as a "control" area for a particular "study" area. For example, for an area where a nuclear power plant is located, a group of 25 adjacent MCDs may be selected as a "study" area; for comparison purposes, a group of 28 adjacent (and different) MCDs may be selected as a "control" area within the Commonwealth

where no nuclear power plants exist. Most frequently used criteria for matching/selecting these areas are population size and race-sex-age composition. Some additional characteristics, such as income, are also used in some instances. The same matching/selecting criteria can be used for an area where potential health risks may exist from chemical environmental hazards and for an area selected as "control" for that study.

While delineation of appropriate geographic location and the size of population to be included ("study" or "control" area) will be dictated by each specific investigation, there must be some general guiding principle that may be applicable to most, if not all, such investigations. For example, when cancer incidence and/or cancer deaths are the primary concern in the study, the total area population ("study" or "control") should be approximately 50,000 within a 10 mile radius, if possible. If, on the other hand, general mortality (all-causes considered together) and/or natality problems (e.g., low birth weight and congenital anomalies) are the primary concern for investigation, the total area population may be smaller, but no less than 30,000 for each geographic area selected, again each within an approximate 10-mile radius, if possible.

B. Data Sources and Health Indices:

(1) Population Data:

Population counts and estimates are essential for the computation of vital statistics rates, which are necessary for comparing vital events between different populations and for ascertaining time trends in the same population. The population data derived from the national decennial censuses represent the most appropriate population statistics needed for detailed demographic and epidemiologic analyses. In addition to population statistics, demographic characteristics of populations under study are also important when morbidity and mortality risks are assessed.

The State Health Data Center has census tapes which contain data for every MCD in the State. This will facilitate obtaining and utilization of census data for the Statewide Epidemiologic Surveillance System. In addition to decennial census data, the Bureau of the Census periodically publishes postcensal estimates of the population (total count without age-sex breakdowns) of each of the MCDs in the State. Other sources of population data for Pennsylvania include the State Data Center, Pennsylvania State University Capitol Campus, which publishes intercensal population estimates for each county, but not for MCDs. The most recent available MCD-based population data come from the 1980 census. Within the next year, the 1990 census, including demographic and socio-economic data by each MCD should be available for the epidemiologic surveillance system.

While census populations (e.g., 1980 and 1990) provide an important component of the necessary data for surveillance purposes, intercensal or postcensal population estimates for individual MCDs do not provide age-sex breakdowns, which are essential for calculation of adjusted (standardized) rates and ratios. As a result, the 1980 census data by age and sex for each MCD will have to be substituted, as best available estimates, for any off-census years until the 1990 census data become available. This can create potential problems for areas where rapid demographic changes have taken place.

(2) Mortality-Natality Data:

The Pennsylvania Department of Health is the custodian of all vital statistics files and records collected under the provisions of Act 66 (1953) of the Pennsylvania General Assembly. The organizational unit within the Department of Health which is responsible for the management of these data is the State Health Data Center.

Birth and death certificates are the single most important source of epidemiologic surveillance. Mortality and natality data, unlike morbidity data, are nearly complete in terms of case ascertainment; thus, such data are of particular value for epidemiologic studies and surveillance.

There are several limitations in the use of existing vital statistics, though. For example, information contained in death certificates is not always complete or accurate. The date of the

initial diagnosis of the disease and environmental-occupational risk exposures to the deceased are not likely to be recorded on death certificates.

Multiple-causes of death are coded on computer tapes in Pennsylvania since 1980, which include both "underlying" and "contributory" causes. Fetal death information in Pennsylvania is limited to products of conception which have achieved sixteen (16) weeks of gestation; thus, a certificate of death is not required (thus, not available) when conception is less than 16 weeks.

The most serious and inherent problem associated with birth and death certificates is the reporting (thus, coding) of mailing address, which may not be the same as place of residence. Specifically, health events tend to be over-reported to boroughs and cities where post offices are located, while such events tend to be under-reported to townships where no post offices are located. This discrepancy can effect the accuracy of the health status indicators, based on the reported (uncorrected) mailing address. However, when several adjacent MCDs, including at least one borough, are considered together, this potentially serious problem may not be a major methodological issue.

(3) Cancer Registry Data:

Cancer registry (incidence) data can provide one of the most reliable and useful statistics for surveillance. The Pennsylvania Cancer Registry was created in 1982 under Act 224

(Cancer Control Act). The Statewide registry provides a basis for ascertaining newly diagnosed cancers, determining cancer incidence in selected geographic locations and assessing secular trends in cancer rates by demographic characteristics. This is important because cancer is "perceived" by the public to be one of the diseases most closely associated with living in the vicinity of potentially high-risk areas, such as nuclear power plants and toxic chemical dump sites. Specific information ascertained routinely for the Cancer Registry include: patient's age, sex, race, residence, marital status, living status, occupation, anatomical (organ) site of cancer, histological diagnosis, stage of the disease, method of diagnosis, and date of death, if applicable.

(4) Minimum Data Set:

Because the surveillance program described in this paper is a "screening" program (a systematic means of signaling the presence of a potentially significant health problem) for selected geographic locations, a set of minimum, but adequate health data is considered to be appropriate. Such data, however, must be routinely available within the State Health Department and should also be updated and analyzed with limited resources.

Another consideration given to the selection of a specific data set is the latency period for the development of diseases, i.e., the time interval from the potential exposure to environmental agents to the onset of diseases and conditions.

For example, pregnancy outcome measures, such as fetal death, low birth weight and congenital anaomalies, are of short latency and can be evaluated within a period of 10 to 12 months post-exposure.

In contrast, cancer, which has a long latency period, cannot be evaluated until "sufficient" time has elapsed. Except for leukemia, which can develop within a few years under certain circumstances, all other forms of cancer usually take 10 to 20 or more years to develop, after exposure to carcinogens.

There is also a time lag from the time/date of cancer diagnosis to the time/date of death. This time lag can also vary substantially from one type of cancer to another. Nevertheless, cancer incidence can be assessed in a historical context, with appropriate adjustment, within a given population or in comparison with other populations. Cancer incidence data should be viewed as an important corollary to the overall health profile which is based upon other diseases and conditions which have a relatively short latency period.

(5) Health Indices:

The short-term health indices for surveillance included in the minimum data set are derived from birth and death (including fetal death) certificates. These are: fetal mortality (16 or more weeks gestation), neonatal mortality (within 28 days of delivery), infant mortality (within 365 days of life), congenital anomalies (recognized at birth), prematurity (delivery with less

than 37 weeks gestation), immaturity (birth weight less than 2,500 grams), and low Apgar scores (less than 7 at one-minute postpartum).

The long-term health indices considered for surveillance would include at least cancer incidence and mortality, all types considered together and certain selected anatomical sites considered separately, depending largely upon the nature of the specific environmental exposure under study. Other chronic diseases and conditions would also be considered, according to the existing knowledge of their etiologic relationship to particular environmental contaminants under study. These chronic diseases and conditions are, therefore, not a part of the routinely analyzed "minimum data set." It should be noted that the presence of an excess cancer problem in a population due to an environmental influence is often correlated to the presence of other health problems, such as congenital anomalies and infant deaths.

C. Data Analysis:

(1) Adjustment (Standardization) of Health Indices:

When baseline data are established and various health status indices computed, a variety of epidemiologic analyses can be conducted. In computing such indices, however, demographic differences in the populations under consideration, which could bias the results, should be adjusted to make comparisons more meaningful and appropriate. Several different methods of

adjustment (standardization) are available, depending upon the prevailing circumstance.

Standardization serves the purpose of refining the comparison by statistically eliminating an "extraneous" source of variation in the data that may seriously affect the results of the analyses. Differences in the age-sex (age-sex-race, if minority populations are sufficiently large) composition are most commonly taken into account when mortality or morbidity rates are standardized. Mother's age and race are often considered for adjustment when perinatal outcomes are compared between populations.

Two established methods, direct and indirect, are available to adjust or standardize mortality and morbidity measures for demographic variables. Each approach has advantages and disadvantages, and the selection of one against another is often dictated by (a) the availability of the necessary data and (b) the size and demographic characteristics of the study and control (standard) populations.

The Direct Method of Standardization (adjustment) is preferable, if the study population is reasonably large and the reliable stratum-specific data (e.g., age-sex specific mortality rates) are available for the study population. Since the age-sex specific rates in each study population are weighted by some common age-sex distribution of a standard (control) population, the resultant standardized rates are mutually comparable. This comparability is an advantage of the direct method. Usually, the

standard (control) population selected for direct adjustment is of reasonably large size and the accurate information on the age-sex distribution must always be available so that the expected number of deaths or diseases can be computed in that standard (control) population. From the expected number, an adjusted rate along with its corresponding Confidence Intervals can be computed for a statistical significance test.

The primary disadvantage of direct method is the potential instability of the stratum-specific rates derived from each study population; this is particularly true when the study population is small in size. Specifically, if the numbers of individuals within separate age groups are small, the death rates (probabilities of dying) would be subject to large fluctuations through the presence or absence of only a few deaths.

In contrast, the Indirect Method of Standardization (adjustment) is preferable, if the study population is relatively small and the stratum-specific data (e.g., age-sex specific cancer incidence rates) are not available for the study population. The standard (control) population selected for indirect adjustment, while reasonably large in size, must also have reliable information on the stratum-specific data (e.g., age-sex specific cancer incidence rates) that can be applied to each study population to arrive at the expected number (e.g., cancer cases) in that study population. When cancer incidence is evaluated, the observed number of cancer cases in the study

population is compared with the expected number based on the age-sex specific cancer incidence rates in the selected standard (control) population. An observed/expected ratio computed in this manner is age-sex adjusted and can be tested for statistical significance by the Poisson model (8).

Because of the efficiency of indirect standardization and its demonstrated utility, one can rely on the computation of standardized ratios for screening large amounts of data and for preliminary analyses. This is a definite advantage for the epidemiologic surveillance being discussed in this report. However, one should be aware that, even though standardized ratios are age-sex adjusted, sets of such ratios are not necessarily mutually comparable, because the age-sex distributions of the various study populations may differ. This would be a disadvantage only if one were interested in comparing one study population directly with another study population. This, however, is not the primary consideration in the epidemiologic surveillance being discussed.

It is also important to note that standardized or adjusted measures have no absolute value in themselves. They are meaningful only in comparison with other similarly computed measures. Since they are useful only for comparison, the commonest application of the procedure is to derive an "observed/expected" ratio, or "standardized mortality ratio" (SMR), to determine, for example, if the cancer mortality rate in a selected population is significantly higher than normal.

(2) Modes of Comparison:

For the purpose of epidemiologic surveillance using the MCD-based minimum data set, three different modes of comparative analyses can be made to detect potentially significant health problems in any geographic area delineated by MCDs. A brief description of each mode follows.

(a) Cross-Sectional Comparison Between Study and Control Areas:

Any study area delineated by MCDs where potential health risks may exist (e.g., EPA/State Superfund sites; nuclear power plant sites) or where alleged health problems (disease clusters) are reported by concerned local residents can be compared with a control area, also delineated by MCDs, where demographic characteristics are similar, but no outstanding environmental health risks are known to exist. This area -to-area comparison is made for a specific time period.

In addition to a specially selected MCD-based control population, health indices for the study population can also be compared with the State average data used as a common control. For each health status measure (e.g., infant mortality, cancer incidence, congenital anomalies, etc.) considered, the "observed" and "expected" values can be derived according to the Indirect Method of Standardization, and then a ratio of "observed" to "expected" computed. For example, from the "observed" and "expected" numbers of deaths, an SMR

(Standardized Mortality Ratio) can be computed for a statistical significance test using the Poisson model.

(b) Temporal Comparison Within Study Population:

Various health indices can be compared from one year to another within the same study population. This comparison is important, and is often more sensitive than that between study and control populations described earlier because health indices are influenced by the existing socio-economic and environmental factors within a given geographic location, and significant changes in such indices are likely to be linked to recognizable changes in the environment. It is important to be aware, however, that health indices in a relatively small population can fluctuate appreciably over time, but still within the limit of random variations, which may give an erroneous impression of "significant" change.

Another advantage of temporal comparison is that health indices within a given geographic area, prior to the operation of a new industrial plant producing environmental health risks, can be directly compared with those after the plant began operation. Logically, this should be a more effective method of assessing the potential impact of this plant upon the adjacent environment.

Historical analyses of this type can provide an additional dimension and have an advantage over cross-sectional comparisons, in that much of the basic environmental characteristics (physical

and socio-economic) tend to remain fairly stable. This would permit relating the changes observed in health status of the local population to recognizable changes in the environment; e.g., such as the establishment of a major industry with potential environmental pollution and unscheduled or accidental releases of toxic substances from an already existing plant in the area.

(c) Temporal -Spatial Comparison:

This would be a combination of cross-sectional and temporal comparisons in which changes over time in health indices within a given study area may be compared with changes, if any, over the same time period in one or more control areas. If similar time trends were observed for both the study and control areas, such changes in health status indicators are more likely due to some common factor(s) rather than due to any particular factor unique to the study area. This analysis would be an important supplement, particularly when the difference between study and control populations and/or changes over time within a given study population is statistically significant at the low or borderline levels.

D. Surveillance Sequence:

Several different steps/tasks are involved in the development and implementation of the Epidemiologic Surveillance System being described.

First, minimum data sets, including both population and health statistics selected for surveillance by all MCDs in Pennsylvania, are compiled on computer tapes for each year as they become available. This compilation is done according to a designated format and layout. Same data set for Pennsylvania, as a whole, is also compiled.

Second, for each of the selected health data items (e.g., mortality), the age-sex specific rates, based on the Pennsylvania population, are computed for a given year and entered onto the computer tapes. These sets of "standard rates" are applicable to any of the MCD groupings selected for specific purposes to compute the "expected" number of health events (e.g., deaths) in the study area being investigated.

Third, each MCD grouping for each of the EPA/State Superfund sites in Pennsylvania can be monitored for possible "abnormalities," by applying the same procedure described above. Specifically, each study area (MCD grouping) and another group of MCDs pre-selected as control will provide the necessary data for a cross-sectional comparison. The difference between "observed" and "expected" numbers can, then, be tested for a statistical significance. Epidemiologic surveillance around Superfund sites will be a routine practice as these sites are pre-determined.

Fourth, each MCD grouping for each of the nuclear power plant sites in Pennsylvania can also be monitored for possible "abnormalities" by going through the same procedure described

above. This will also be a routine practice as such sites are already identified.

Fifth, if the results of the routine screening analyses described above are justified, special in-depth studies may be conducted. This justification would require more than what was found in the surveillance process in order to conclude that the observed difference is "real". If "real", then epidemiologic studies, including case-control design, retrospective cohort design, and/or clinical observations, would be conducted to determine possible causes(s) of the observed difference.

Sixth, for requests investigating reported "disease clusters" within the Commonwealth, regardless of the recognition of potential environmental hazards, the study area covering a certain number of adjacent MCDs will be compared with another group of control MCDs for a possible, significant difference in terms of cancer incidence, mortality or any other health indices under consideration. For "disease cluster" analyses, control MCDs will be selected only when such a study is commenced and only for that particular study area.

E. Limitations and Qualifications:

Regardless of the methods of standardization (adjustment) used, a study population (a group of MCDs) can be compared directly with a control (standard) population (a group of MCDs) selected on a one-to-one basis. However, for practical reasons,

the Indirect Method of Standardization, using the State average data, would be more useful for routine surveillance.

One particular aspect of the statistical significance tests, which is often misunderstood, needs to be clarified. When numerous statistical comparisons are made, as is the case with the Epidemiologic Surveillance System being described, one should be aware that one or two statistically significant findings are not at all unusual and such findings, by themselves, do not necessarily indicate the existence of a "health problem" in the area. In fact, it would be possible (or even likely) to see some statistically significant differences/changes in health indices, which may not necessarily be of epidemiologic importance. For this reason, the results of routine surveillance should be viewed with caution.

One potential limitation of the MCD-based epidemiologic surveillance system in Pennsylvania is that the place of residence is not always consistent with the mailing address. This problem can arise when a MCD is miscoded and, consequently, the observed number of heatlh events (e.g., deaths) for that particular MCD is either inflated (if, borough or city) or deflated (if, township). However when several adjacent MCDs are combined, as described for the surveillance system, this potential problem should be minimized.

V. Operational Procedures:

The Division of Epidemiology Research and the State Health Data Center have already developed the mechanism to routinely compile available vital statistics and population data on computer tapes for the epidemiologic surveillance around nuclear power plants in Pennsylvania. Compilation of a minimum data set for five nuclear power plant sites and selected control areas for the 1979-1983 period has been completed and the computerized procedures have been tested. This should be re-evaluated and appropriate adjustments made, as and when necessary.

A similar procedure can be developed for other geographic areas delineated by MCDs including any of the EPA/State-designated Superfund sites and any other specified communities where "disease clusters" are reported, regardless of environmental implications.

In terms of Automated Systems, which should be applicable to these surveillance purposes, the baseline data, including selected minimum health data and appropriate population statistics for a given year, need to be entered onto computer tapes. Such baseline data sets are subject to periodic update with new, more recent data when they become available. Newly created computer tapes for surveillance contain only the "minimum" data elements for each MCD in Pennsylvania. Direct access to these tapes is limited to authorized members of the Division of Epidemiology Research.

Once data elements for a given year are entered onto computer tapes for each MCD in the Commonwealth, MCDs can be grouped into a "study" or "control" area, depending upon the specific purpose. For a surveillance purpose, such as for the nuclear power plant sites or for the EPA-State Superfund sites, comparable health and population data for Pennsylvania, as a whole, may be used as the common "control" or "standard."

When direct comparison is made between the "study" and "control" areas, each based on selected MCDs, or when the "study" area is compared with Pennsylvania, as a whole, used as the common "control", some of the routinely available demographic confounders must be taken into account. As indicated earlier, computerized system must be developed in order to adjust for such differences between the areas being compared.

For mortality analyses, both underlying and contributory causes should be considered and analysis made for all-causes combined and selected causes separately. At least age at death, sex, and race (if data permit) should be taken into account in the comparisons. In contrast, for natality analyses, including congenital anomalies, low birth weight, premature delivery, and low Apgar scores, at least sex/race of the newborn and mother's age and level of education should be taken into account in the comparisons. The Indirect Method of Standardization would be the choice for these adjustments.

Some appropriate computer program(s) for standardization need to be developed to adjust for demographic and/or social

variables simultaneously that can influence the designated health indices. (This is a different technical issue which should be addressed under separate cover and is not addressed in this report.)

VI. Significance and Public Health Implications:

The number of requests for epidemiologic assessments or studies from local citizens, who are concerned with "perceived" health problems in their respective communities, has been increasing during the past decade. Many of these concerns are related to certain geographic locations with known potential environmental risks, such as radiation from nuclear power plants or toxic substances/chemicals from industrial waste dump sites.

In a large industrial state, such as Pennsylvania, the State Health Department should respond, in a timely manner, to such inquires and take whatever necessary actions promptly. Presently, the State Health Department, through several Divisions, responds to the various requests for an epidemiologic assessment of perceived health problems on an ad hoc basis.

While the standardization processes have been computerized for certain specific purposes (e.g., cancer data), a uniform operational mechanism, serving multiple surveillance purposes, has not yet been developed.

As stated earlier, the Epidemiologic Surveillance System described here is intended to be an automated system of surveillance within the State of Pennsylvania that can be applied

to all (except infectious-communicable disease) epidemiologic assessments under a variety of different circumstances. The systems development and computerization process, as well as initial data entry and update, would require a considerable undertaking for the State Health Department. However, a multi-purpose surveillance system of this nature is an essential component of contemporary public health as it provides the basis for almost all program planning and evaluation. Without such a system, the State Health Department would remain capable of limited responses to the expressed needs of concerned citizens and it's public health practice compromised of the benefit of the available epidemiologic knowledge.

In terms of a broad public health mandate, this surveillance system also has psycho-educational implication. That is, when such a system has been implemented by the State, both the general public and concerned industries can recognize and understand its value and may prevail upon the services it provides. Furthermore, the surveillance system, as conceived and developed, is capable of leading quickly into an in-depth epidemiologic study, if and when indicated, such as another nuclear accident or any other industrial accidents creating serious environmental hazards.

Another important policy implication is that the epidemiologic surveillance system being described is consistent with the overall Commonwealth's goals/objectives in which other State and Federal agencies also participate. For example, the

epidemiologic surveillance around nuclear power plants conducted by the State Health Department is closely linked to the roles of the Pennsylvania Emergency Management Agency and the Pennsylvania Departments of Agriculture and Environmental Resources. Likewise, the epidemiologic surveillance around toxic chemical dump sites conducted by the Pennsylvania Department of Health has its conterparts in both Federal (Environmental Protection Agency and Agency for Toxic Substance and Disease Registry) and State (Department of Environmental Resources) agencies. Thus, an epidemiologic surveillance system of this nature will help enhance (promote) the growing need for a Federal-State partnership in vital public health mandates.

REFERENCES

1. Tokuhata, G.K. and Smith, M.W. History of Health Studies Around Nuclear Facilities: A Methodological Consideration. Environ. Res. 25:75-85, 1981.

2. Lechat, M.F. Disasters and Pulbic Health. World Health Organization 57:11-17, 1979.

3. Pell, S., O'Berg, M.T., and Karrh, B.W. Cancer Epidemiologic Surveillance in the DuPont Company. J. Occup. Med. 20, Nov. 1978.

4. DeGroot, M.H. Statistical Studies of the Effect of Low Level Radiation from Nuclear Reactors on Human Health. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. Vol. VI:223-234, 1972.

5. Patrick, C.H. Trends in Public Health in the Population Near Nuclear Facilities: A Critical Assessment. Nuc. Safety 18:647-662, 1977.

6. Tompkins, E.A., Hamilton, P.M., and Hoffman, D.A. Infant Mortality Around Three Nuclear Power Reactors. In: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability. Vol. VI:279-290, 1972.

7. Grahn, D. Analysis of Population, Birth, and Death Statistics in the Counties Surrounding the Big Rock Point Nuclear Power Station, Charlevoix County, Michigan. Argonne National Laboratory, Argonne, IL, 1975.

8. Bailor, J.C. III and Ederer, F. Significant Factors for the Ratio of a Poisson Variable to Its Expectation. Biometrics 20:639-643, 1964.